

KLUSTER K-MEANS DATA MAHASISWA BARU TERHADAP PROGRAM STUDI YANG DIPILIH

Citra Arum Sari dan Dwi Sukma D
Program Studi Teknik Industri, FTI-UPN Jatim

ABSTRAK

Besarnya peminat dari setiap program studi di UPNV Jawa Timur berbeda – beda. Berdasarkan hal tersebut, maka dalam penelitian ini bertujuan untuk mengelompokkan peminat program studi untuk memberikan informasi kepada pihak Universitas sehingga dapat memprioritaskan program studi peminat sedikit.

Obyek penelitian ini adalah peminat mahasiswa masuk ke program studi UPNV Jatim Tahun Akademi 2013/2014. Variabel bebas dalam penelitian ini adalah asal SMA, pendapatan orang tua dan asal kota. Variabel terikat adalah pengelompokan peminat program studi. Pengumpulan data didapat dari Biro Admik UPNV Jatim yaitu data penerimaan mahasiswa yang berminat masuk ke program studi pada TA.2013/2014. Pengolahan data menggunakan teknik data mining dan menggunakan clustering K-Means untuk mengelompokkan program studi dalam beberapa kelas berdasarkan kemiripan data.

Berdasarkan hasil clustering dengan metode algoritma K-Means maka dapat disimpulkan bahwa terdapat 3 kluster, dimana kluster 1 merupakan program studi dengan sedikit peminat sebanyak 10 program studi, kluster 2 banyak peminat sebanyak 5 program studi dan kluster 3 cukup peminat sebanyak 4 program studi.

Kata Kunci : Program Studi, Data Mining, Clustering K-Means

ABSTRACT

The amount of interest from each subject at UPNV East Java is different. Based on this, the research aims to group study program to provide information to the University so as to prioritize the program of study with little interest.

The data used is the data the student population in 2013/2014 UPNV East Java from each subject. The independent variable in this study is the origin of high school, parental income and home town. Data obtained from the Bureau Admik UPNV East Java, using data mining techniques, while processing the data using K-Means clustering to group courses in several classes based on the similarity of data.

Based on the results of the clustering method K-Means algorithm, it can be concluded that there are three clusters, where cluster 1 is a program of study with little interest in as many as 10 courses, cluster 2 is many enthusiasts as many as 5 courses and clusters 3 is quite interested by 4 courses.

Keyword : Courses, Data Mining, Clustering K-Means

PENDAHULUAN

Universitas Pembangunan Nasional “Veteran” Jawa Timur, disingkat UPNV Jatim atau UPN VJT berlokasi di Surabaya merupakan Perguruan Tinggi Swasta di Indonesia yang berdiri sejak 5 Juli 1959. UPNV Jawa Timur hingga tahun 2013, memiliki 6 Fakultas dengan 19 Program Studi (Progdi) sarjana, yaitu Fakultas Ekonomi Bisnis dengan 3 Program Studi, yaitu Progdi Ilmu Ekonomi dan Pembangunan, Progdi Akuntansi dan Progdi Manajemen, Fakultas Petanian dengan 2 Program Studi, yaitu Progdi Agroteknologi dan Progdi Agribisnis, Fakultas Teknologi Industri dengan 5 Program Studi, yaitu Progdi Teknik Kimia, Progdi Teknik Industri, Progdi Teknologi Pangan, Progdi Teknik Informatika dan Progdi Sistem Informasi, Fakultas Ilmu Sosial dan Ilmu Politik dengan 4 Program Studi, yaitu Progdi Administrasi Negara, Progdi Administrasi Bisnis, Progdi Ilmu Komunikasi dan Progdi Hubungan Internasional, Fakultas Teknik Sipil dan Perencanaan dengan 4 Program Studi, yaitu Progdi Teknik Sipil, Progdi Teknik Arsitektur, Progdi Teknik Lingkungan dan Progdi Desain Komunikasi Visual, serta Fakultas Hukum dengan 1 Program Studi yaitu Progdi Ilmu Hukum.

Jumlah mahasiswa baru di Universitas Pembangunan Nasional “Veteran” Jawa Timur mengalami pertumbuhan dan perkembangan yang cukup signifikan. Jumlah mahasiswa baru dapat diketahui dari setiap program studi yang ada, dimana ada yang banyak peminat dan kurang peminat. Besarnya peminat dari setiap program studi dapat dipengaruhi oleh asal kota, pendapatan orang tua, asal wilayah dan lain – lain. Berdasarkan hal tersebut, maka dalam penelitian ini akan mengelompokkan program studi berdasarkan data mahasiswa yang ada dengan tujuan memberikan informasi kepada pihak Universitas untuk memprioritaskan program studi yang memiliki sedikit peminat.

Untuk metode yang akan digunakan dalam penelitian ini adalah teknik data mining. Data mining berperan sebagai pencarian informasi yang berharga dari basis data yang sangat besar. Data mining adalah suatu proses dalam menemukan berbagai model, ringkasan data dan nilai – nilai yang berharga dari sekumpulan data. Pada penelitian ini, metode data mining digunakan untuk proses pencarian informasi dari data mahasiswa baru semester 1 UPNV Jawa Timur. Selain data mining juga menggunakan teknik *clustering K-Means*, dimana *clustering K-Means* digunakan untuk mengelompokkan program studi berdasarkan jarak minimum setiap data ke kluster.

Dengan demikian, diharapkan dapat memberikan informasi yang bermanfaat bagi pihak Universitas dalam melakukan promosi mengenai program studi sarjana yang ada di Universitas Pembangunan Nasional “Veteran” Jawa Timur ini.

Tinjauan Pustaka

Data Mining

Data *mining* adalah analisis otomatis dari data yang berjumlah besar atau kompleks dengan tujuan untuk menemukan pola atau kecenderungan yang penting yang biasanya tidak disadari keberadaannya (Moertini, 2002).

Hal-hal penting yang terkait dengan data *mining* adalah (Luthfi & Kusrini, 2009):

1. Data *mining* merupakan suatu proses otomatis terhadap data yang sudah ada.
2. Data yang akan diproses berupa data yang sangat besar.
3. Tujuan data *mining* adalah mendapatkan hubungan atau pola yang mungkin memberikan indikasi yang bermanfaat.

Data *mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu (Larose, 2005):

1. Deskripsi
Terkadang penelitian analisis secara sederhana ingin mencoba mencari cara untuk menggambarkan pola dan kecenderungan yang terdapat dalam data.
2. Estimasi
Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori.
3. Prediksi
Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada di masa datang.
4. Klasifikasi
Dalam klasifikasi, terdapat target variabel kategori.
5. Pengklusteran
Pengklusteran merupakan pengelompokan *record*, pengamatan, atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan.
6. Asosiasi
Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu.

Clustering

Baskoro (2010) menyatakan bahwa *Clustering* atau *clusterisasi* adalah salah satu alat bantu pada *data mining* yang bertujuan mengelompokkan obyek-obyek ke dalam *cluster-cluster*. *Cluster* adalah sekelompok atau sekumpulan obyek-obyek data yang *similar* satu sama lain dalam cluster yang sama dan *dissimilar* terhadap obyek-obyek yang berbeda *cluster*. Obyek akan dikelompokkan ke dalam satu atau lebih *cluster* sehingga obyek-obyek yang berada dalam satu *cluster* akan mempunyai kesamaan yang tinggi antara satu dengan lainnya. Obyek-obyek dikelompokkan berdasarkan prinsip memaksimalkan kesamaan obyek pada *cluster* yang sama dan memaksimalkan ketidaksamaan pada *cluster* yang berbeda. Kesamaan obyek biasanya diperoleh dari nilai-nilai atribut yang menjelaskan obyek data, sedangkan obyek-obyek data biasanya direpresentasikan sebagai sebuah titik dalam ruang multidimensi.

Algoritma K-Means

Menurut Widyawati (2010), algoritma *k-means* merupakan algoritma yang membutuhkan parameter input sebanyak *k* dan membagi sekumpulan *n* objek kedalam *k cluster* sehingga tingkat kemiripan antar anggota dalam satu *cluster* tinggi sedangkan tingkat kemiripan dengan anggota pada *cluster* lain sangat rendah. Kemiripan anggota terhadap *cluster* diukur dengan kedekatan objek terhadap nilai *mean* pada *cluster* atau dapat disebut sebagai *centroid cluster* atau pusat massa.

Dalam penggunaan algoritma *K-Means* memiliki beberapa kelemahan dan kelebihan, yaitu :

A. Kelebihan K-Means :

1. Selalu konvergen atau mampu melakukan klusterisasi.
2. Tidak membutuhkan operasi matematis yang rumit, bisa dibilang operasinya sederhana.
3. Beban komputasi relatif lebih ringan, sehingga klusterisasi bisa dilakukan dengan cepat walaupun relatif tergantung pada banyak jumlah data dan jumlah *cluster* yang ingin dicapai.

B. Kekurangan K-Means :

1. *K-Means* hanya bisa digunakan untuk data yang atributnya bernilai numerik. Jumlah *cluster* sebanyak *K*, harus ditentukan sebelum dilakukan perhitungan.

2. Nilai *centroids* yang diberikan di awal bisa mempengaruhi hasil klusterisasi apabila nilainya berbeda (sensitif terhadap nilai *centroids* awal).
3. Solusi *cluster* yang dihasilkan hanya bersifat *local optima*, sehingga kita tidak tahu apakah itu sudah merupakan konfigurasi optimal atau belum.
4. Tergantung pada *mean* (rata – rata).
5. Algoritma.

METODE PENELITIAN

Pengamatan Obyek

Pengamatan obyek ini dilakukan di UPNV Jawa Timur. Waktu penelitian dilaksanakan pada bulan September 2012 hingga data yang dibutuhkan tercukupi.

Langkah-Langkah Penelitian dan Pemecahan Masalah

Adapun tahapan penelitian yang akan dilakukan dalam proses penelitian ini adalah dengan mengumpulkan dan mempelajari literatur yang berkaitan dengan konsep Data Mining *clustering*, yang menggunakan algoritma *K-Means*. Sumber literatur berupa buku teks, *paper*, jurnal, karya ilmiah, dan situs-situs penunjang, mengumpulkan data dan informasi yang didapat dari Biro Admik UPNV Jatim yang meliputi asal kota, pendapatan orang tua, jenis SMA dan program studi. *Clustering* menggunakan Algoritma *K-Means* menentukan *k* sebagai jumlah *cluster* yang ingin dibentuk, membangkitkan *k centroids* (titik pusat *cluster*) awal secara random, menghitung masing-masing jarak setiap data ke masing-masing *centroids*, memilih *centroids* yang terdekat, menentukan posisi *centroids* baru dengan cara menghitung nilai rata-rata dari data-data yang memilih pada *centroids* yang sama, kembali ke langkah 3 jika posisi *centroids* baru dengan *centroids* lama tidak sama, menganalisa hasil yang sudah diperoleh pada proses *clustering*. Setelah pengolahan dan analisa data, maka langkah selanjutnya adalah menarik kesimpulan, kemudian diberikan juga saran sebagai rekomendasi yang dapat memberikan manfaat bagi peneliti maupun pihak UPNV Jawa Timur.

HASIL DAN PEMBAHASAN

Pengumpulan Data

Data yang digunakan untuk pengolahan data nantinya yaitu data sekunder yang diperoleh melalui Biro Admik UPNV Jawa Timur.

Tabel 1. Data Mahasiswa setisp Program Studi

Prodi	SMA	SMK	Pendapatan orang tua			Wilayah		
			I	II	III	I	II	III
11	51	16	15	14	38	66	1	0
12	258	56	36	94	184	308	1	5
13	281	54	47	110	178	329	2	4
24	47	7	27	15	12	53	1	0
25	19	45	34	13	17	60	4	0
31	91	8	29	37	33	96	2	1
32	158	86	98	117	29	239	5	0
33	36	13	23	13	13	49	0	0
34	155	87	58	81	103	235	4	3
35	51	40	33	38	20	89	1	1
41	75	12	32	29	26	84	2	1
42	93	16	35	58	16	108	1	0
43	195	81	63	121	92	270	5	1

44	58	9	20	23	24	61	4	2
51	37	14	29	15	7	48	2	1
52	75	9	39	21	24	79	4	1
53	104	23	43	67	17	122	3	2
54	40	23	27	20	16	62	0	1
71	120	8	42	59	27	124	3	1

Sumber : Data Biro Admik UPNV Jawa Timur

Pengolahan Data

Berikut akan dijelaskan mengenai langkah – langkah pengolahan data dari penelitian yang dilakukan menggunakan Algoritma *K-Means* dengan bantuan *software Matlab* :

1. Menentukan jumlah *cluster* yang akan dibuat. Dalam penelitian ini jumlah *cluster* yang akan dibuat yaitu sebanyak 3.
2. Menentukan nilai pusat *cluster*. Dalam penelitian ini nilai pusat awal ditentukan secara *random* dan didapat nilai pusat dari setiap *cluster*.

Tabel 2.Nilai Pusat Awal

Cluster	SMA	SMK	Pend I	Pend II	Pend III	Wilayah I	Wilayah II	Wilayah III
1	47	7	27	15	12	53	1	0
2	155	87	79	132	31	235	4	3
3	120	8	42	59	27	124	3	1

Sumber: data diolah

3. Menghitung jarak antara *centroid* dengan masing-masing data. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling dekat antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana, hasil dapat dilihat pada tabel berikut :

Tabel 3.Jarak Setiap Program Studi dengan Pusat Kluster Iterasi 1

Jarak 1	Jarak 2	Jarak 3
1087	62505	11069
147878	43614	81134
169731	35187	91147
0	67960	12762
2364	67276	17948
4717	41195	2315
68908	626	27262
435	70819	15049
67960	0	26506
2911	45957	7755
2261	47359	4541
7151	32669	1225
93418	6964	40784
363	60691	9451
541	69511	14731
2230	48886	4902
11259	25185	651
597	62001	12107
12762	26506	0

Sumber: Data diolah

4. Memasukkan data ke dalam setiap kluster. Setelah menghitung jarak setiap data, langkah selanjutnya adalah mengelompokkan data berdasarkan hasil minimum jarak.

Berdasarkan hasil pengolahan data, iterasi *clustering* dilakukan sebanyak 3 kali sampai didapatkan nilai pusat kluster dan pengelompokkan data pada setiap kluster tidak berpindah.

Tabel 4 Hasil Pengolahan Data Program Studi

Cluster	Program Studi	Keterangan
1	Ekonomi Pembangunan Agribisnis Agroteknologi Teknik Pangan Sistem Informasi Ilmu Admin Negara Hubungan Internasional Teknik Arsitektur Teknik Lingkungan Desain Komunikasi Visual	Mahasiswa yang berasal dari SMA 489 mahasiswa, SMK 188 mahasiswa, pendapatan orang tua I 245 mahasiswa, pendapatan orang tua II 235 mahasiswa, pendapatan orang tua III 197 mahasiswa, Wilayah I 651 mahasiswa, Wilayah II 19 mahasiswa, wilayah III 7 mahasiswa.
2	Manajemen Akuntansi Teknik Industri Teknik Informatika Ilmu Komunikasi	Mahasiswa yang berasal dari SMA 1047 mahasiswa, SMK 364 mahasiswa, pendapatan orang tua I 302 mahasiswa, pendapatan orang tua II 523 mahasiswa, pendapatan orang tua III 586 mahasiswa, Wilayah I 1381 mahasiswa, Wilayah II 17 mahasiswa, wilayah III 13 mahasiswa.
3	Teknik Kimia Ilmu Admin Niaga Teknik Sipil Ilmu Hukum	Mahasiswa yang berasal dari SMA 408 mahasiswa, SMK 55 mahasiswa, pendapatan orang tua I 149 mahasiswa, pendapatan orang tua II 221 mahasiswa, pendapatan orang tua III 93 mahasiswa, Wilayah I 450 mahasiswa, Wilayah II 9 mahasiswa, wilayah III 4 mahasiswa.

Sumber: data diolah

Pembahasan

Dari hasil *clustering K-means* dengan menggunakan *Matlab* diperoleh 3 *cluster*. Yang termasuk dalam kluster 1 terdapat 10 anggota dari program studi berdasarkan data mahasiswa yaitu kelompok program studi Ekonomi Pembangunan, Agribisnis, Agroteknologi, Teknik Pangan, Sistem Informasi, Ilmu Admin Negara, Hubungan Internasional, Teknik Arsitektur, Teknik Lingkungan dan Desain Komunikasi Visual.

Termasuk dalam kluster 2 terdapat 5 anggota dari dari program studi berdasarkan data mahasiswa yaitu kelompok program studi Manajemen, Akuntansi, Teknik Industri, Teknik Informatika dan Ilmu Komunikasi.

Dan yang termasuk dalam kluster 3 terdapat 4 anggota anggota dari dari program studi berdasarkan data mahasiswa yaitu kelompok program studi Teknik Kimia, Ilmu Admin Negara, Teknik Sipil dan Ilmu Hukum.

Berdasarkan hasil analisa kluster di atas, maka dapat diketahui bahwa mahasiswa yang berasal dari SMA lebih banyak dibandingkan dengan mahasiswa dari SMK. Hal ini karena SMA lebih menekankan pada penguasaan ilmu pengetahuan sebagai bekal untuk melanjutkan pendidikan yang lebih tinggi, sedangkan SMK lebih menekankan pada

penguasaan ketrampilan sebagai bekal untuk memasuki dunia kerja. Selain itu, faktor keluarga maupun kemampuan ekonomi keluarga juga sangat berpengaruh terhadap mahasiswa tersebut untuk melanjutkan pendidikan yang lebih tinggi.

Berdasarkan pendapatan orang tua dapat diketahui bahwa mahasiswa yang kuliah berdasarkan pendapatan orang tua II lebih banyak dibandingkan dengan mahasiswa yang kuliah berdasarkan pendapatan orang tua I dan pendapatan orang tua III dari semua program yang ada. Hal ini dapat dikarenakan mahasiswa yang berkuliah disesuaikan dengan kemampuan kondisi ekonomi keluarga, keinginan orang tua ataupun keinginan mahasiswa itu sendiri.

Berdasarkan wilayah atau asal kota mahasiswa dapat diketahui bahwa mahasiswa yang berasal dari wilayah I lebih banyak daripada mahasiswa yang berasal dari wilayah II dan III. Hal ini dapat dikarenakan kurangnya promosi atau sosialisasi yang dilakukan pada wilayah – wilayah I dan II.

KESIMPULAN

Berdasarkan hasil penelitian di atas, maka dapat disimpulkan bahwa pada kluster 1 adalah kelompok program studi dengan cukup peminat yaitu program studi Ekonomi Pembangunan, Agribisnis, Agroteknologi, Teknik Pangan, Sistem Informasi, Ilmu Admin Negara, Hubungan Internasional, Teknik Arsitektur, Teknik Lingkungan dan Desain Komunikasi Visual. Pada kluster 2 adalah kelompok program studi dengan banyak peminat yaitu program studi Manajemen, Akuntansi, Teknik Industri, Teknik Informatika dan Ilmu Komunikasi. Dan kluster 3 adalah kelompok program studi dengan cukup peminat yaitu program studi Teknik Kimia, Ilmu Admin Negara, Teknik Sipil dan Ilmu Hukum. Maka, kelompok program studi yang perlu mendapatkan prioritas utama dari pihak Universitas adalah kelompok program studi yang berada pada kluster 1.

DAFTAR PUSTAKA

- Albar, Ismail & Fibriyanti.2010. **Identifikasi Dengan Menggunakan Algoritma K-Means Pada Plat Kendaraan.** Jurnal.Padang: Jurusan Teknik Elektro.Politeknik Negeri Padang.
- Budiman, Irwan.2012.**Data Clustering Menggunakan Metodologi Crisp-Dm Untuk Pengenalan Pola Proporsi Pelaksanaan Tridharma.**Semarang:Universitas Diponegoro.
- Firdausi, Nabila. Dkk. (2011).**Analisis Financial Distress Dengan Pendekatan Data Mining Pada Industri Manufaktur Go-Public Di Indonesia.** Surabaya : Jurusan Teknik Industri, Institut Teknologi Sepuluh Nopember.
- Martiana S.Kom,M.Kom, Entin.Dkk.2010.**Penggunaan Metode Pengklasteran Untuk Menentukan Bidang Tugas Akhir Mahasiswa Teknik Informatika Pens Berdasarkan Nilai.**Jurnal.Surabaya:Jurusan Teknik Informatika .Politeknik Elektronika Negeri Surabaya.Institut Teknologi Sepuluh Nopember.
- Narwati.2010.**Pengelompokan Mahasiswa Menggunakan Algoritma K-Means.**Jurnal.Fakultas Teknologi Informasi.
- Nur Wahyudi,Eko.Dkk.2011.**Analisa Profil Data Mahasiswa Baru terhadap Program Studi yang dipilih di Perguruan Tinggi Swasta Jawa Tengah dengan Menggunakan Teknik Data Mining.**Jurnal.Program Studi Sistem Informasi.Universitas Stikubank.

- Nur Wahyudi, Eko.Dkk.2010.**Analisa Profil Data Mahasiswa Baru Universitas Stikubank (UNISBANK) Semarang Tahun 2005-2010 Dengan Teknik Data Mining.**
- Nur Wahyudi, Eko.2013.**Teknik Klasifikasi untuk Melihat Kecenderungan Calon Mahasiswa Baru dalam Memilih Jenjang Pendidikan Program Studi di Perguruan Tinggi.**Jurnal.Semarang:Program Studi Sistem Informasi.Universitas Stikubank.
- Oscar Ong, Johan.2013.**Implementasi Algoritma *K-Means Clustering* Untuk Menentukan Strategi *Marketing President University*.**Jurnal.Bekasi: Program Studi Teknik Industri.President University.
- Wakhidah, Nur.***Clustering Menggunakan K-Means Algorithm (K-Means Algorithm Clustering)*.**Jurnal.Semarang: Fakultas Teknologi Informasi dan Komunikasi.Universitas Semarang.